On the geometry of learning neural quantum states [PRR **2**, 023232 (2020)]

Chae-Yeun Park

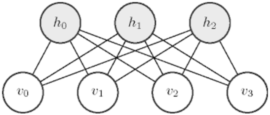
Institute for theoretical physics, University of Cologne

Jul 06, 2020

Natural gradient descent in ML and vQMC

$$\theta_{t+1} = \theta_t - S^{-1}\vec{f} \tag{1}$$

In Machine Learning (ML), $S_{ij} = \langle \frac{\partial p_\theta(x)}{\partial \theta_i} \frac{\partial p_\theta(x)}{\partial \theta_j} \rangle - \frac{\partial p_\theta(x)}{\partial \theta_i} \frac{\partial p_\theta(x)}{\partial \theta_j}$ is the Fisher matrix and $\vec{f}$ the gradient of negative logarithmic likelihood $\sum_x -q(x) \log p_\theta(x)$. In vQMC, we instead have $S_{ij} = \langle \frac{\partial \psi_\theta(x)^*}{\partial \theta_i^*} \frac{\partial \psi_\theta(x)}{\partial \theta_j} \rangle - \langle \frac{\partial \psi_\theta(x)^*}{\partial \theta_i^*} \rangle \langle \frac{\partial \psi_\theta(x)}{\partial \theta_j} \rangle$ which is the quantum Fisher matrix and $f_i = \frac{\partial \langle H \rangle}{\partial \theta_i^*}$.



We here use the restricted Boltzmann machine (RBM) wavefunction that is given by

$$\psi_\theta(x) = \sum_h e^{a \cdot x + b \cdot y + x^T W y}. \tag{2}$$

Dynamics of the quantum Fisher matrix in learning process

We use the complex-valued RBM to solve the transverse field Ising model given by $H = -\sum_i \sigma_z^i \sigma_z^{i+1} + h\sigma_z^i$.

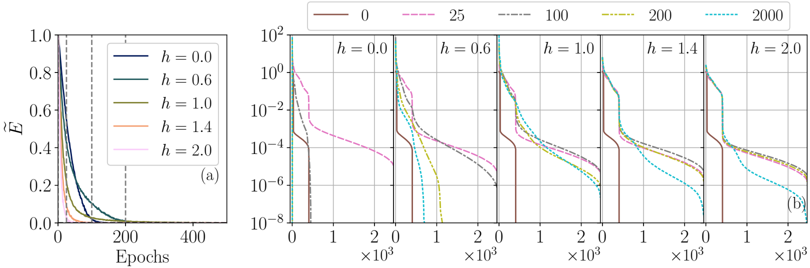


Figure 1: (a) Normalized energy $(\langle E \rangle - E_{\mathrm{ed}})/(E_{\mathrm{ini}} - E_{\mathrm{ed}})$ as a function of epochs. The size of system $N = 28$ is used. (b) Spectra of the quantum Fisher matrix at different epochs.

Initial dynamics pushes up the eigenvalues and is universal. The converged spectra depend on the phases.

Phase dependent convergence property

- In the ferromagnetic phase ($h < 1$), the quantum Fisher matrix is singular at convergence. This is from the fact that the distribution $|\psi_\theta(x)|^2$ is peaked.
- In the critical case ($h = 1$), the spectrum becomes smooth. This resembles a classical ML model when it learns image data. Long-range correlation of the underlying distribution as well as the learning process might be a reason.
- In the paramagnetic case ($h > 1$), the spectrum preserves one of the kinks (located at $N(N+1)/2$) in the initial spectrum. This step-wise behavior is a result of a symmetry of the model parameters.

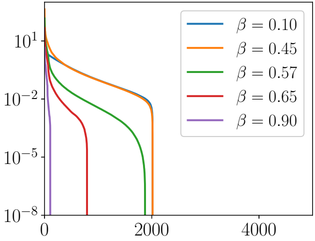


Figure 2: Spectra of the quantum Fisher matrix when the RBM represents coherent Gibbs states of the 2D classical Ising model exactly. In the ferromagnetic case $\beta > \beta_c$, the spectrum becomes singular. Kink appears at $N(N+1)$ in the paramagnetic case.

Remarks and further considerations

- Quantum Fisher information matrix reveals the information of the encoded state.
- In ML, the spectrum of the Fisher matrix is usually smooth and variants of the first-order methods (e.g. RMSProp, AdaGrad, Adam) work incredibly well. In vQMC, the first-order methods often fails and singularity of the quantum Fisher matrix can be a reason.
- In ML, flat minima are often considered as a key to the generalization. In vQMC, this may be interpreted as robustness in representation.